Navy Personnel Research and Development Center
San Diego, California 92152-7250 TN-94-18 May 1994

AD-A280 023

Comparison of a Back Propagation Artificial Neural Network Model With a Linear Regression Model for Personnel Selection

DTIC
ELECTE
JUN 07 1994
S G D

C. A. Wilkins
W. A. Sands

DTIC QUALITY INSPECTED 2

94-17017

94 6 6 089

Approved for public release; distribution is unlimited.

NPRDC-TN-94-18 May 1994

Comparison of a Back Propagation Artificial Neural Network Model With a Linear Regression Model for Personnel Selection

C. A. Wilkins
W. A. Sands

Reviewed by
Kathleen E. Moreno

Approved and released by
Frank L. Vicino
Director, Personnel Systems Department (Acting)

Approved for public release;
distribution is unlimited.

Accesion For		
NTIS CRA&I		X
DTIC TAB		☐
Unannounced		☐
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and / or Special	
A-1		

Navy Personnel Research and Development Center
San Diego, CA 92152-7250

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
May 1994

3. REPORT TYPE AND DATE COVERED
Interim—1 Oct 90-30 Sep 91

4. TITLE AND SUBTITLE
Comparison of a Back Propagation Artificial Neural Network Model With a Linear Regression Model for Personnel Selection

5. FUNDING NUMBERS
Program Element 0601152N
Work Unit 0601152N.R0001.10

6. AUTHOR(S)
C. A. Wilkins, W. A. Sands

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Navy Personnel Research and Development Center
San Diego, CA 92152-7250

8. PERFORMING ORGANIZATION REPORT NUMBER
NPRDC-TN-94-18

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Chief of Naval Research (Code 342)
800 North Quincy Street
Arlington, VA 22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Functional Area: Personnel
Product Line: Printed Testing
Effort: Personnel Screening

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE
A

13. ABSTRACT *(Maximum 200 words)*

This exploratory research examined the usefulness of an Artificial Neural Network—Back Propagation (ANN-BP) model for predicting dichotomous criteria in personnel selection. The results are quite encouraging. Under a wide variety of circumstances, the ANN-BP model was superior to the Ordinary Least-Squares—Linear Regression (OLS-LR) model in predicting curvilinear relationships. Only once was OLS-LR superior. When the underlying relationship was linear, there were no significant differences. This is somewhat remarkable, given that the OLS-LR model was designed to perform optimally in the linear case.

14. SUBJECT TERMS
Artificial neural networks, back propagation, personnel selection

15. NUMBER OF PAGES
17

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

Foreword

This research, conducted under the 6.1 Independent Laboratory Research program, explored the feasibility of applying an Artificial Neural Network—Back Propagation model to personnel selection. Currently, Ordinary Least-Squares Linear Regression is used for many selection problems. The research compared the advantages and drawbacks of using these two methods for personnel selection.

This investigation was sponsored by the Chief of Naval Research (Code 342) and funded by Program Element 0601152N, Work Unit 0601152N.R0001.10. Results are intended for use by the research community.

FRANK L. VINCINO
Director, Personnel Systems Department (Acting)

Summary

Background and Problem

Many important criteria used in military personnel selection are dichotomous (or artificially dichotomized) variables. One important example is the successful completion of first tour of obligated service versus premature attrition. Often, prediction models that exploit these criteria are developed using Ordinary Least-Squares—Linear Regression (OLS-LR) techniques. Such techniques are easy to use but, unfortunately, they restrict the range of models which can be developed. For example, potential predictors that are related to the criterion in a nonlinear fashion, are often rejected by a OLS-LR model.

One course of action would be to apply new mathematical techniques which take into account nonlinear predictor-criterion relationships without requiring extensive knowledge of the underlying relationships. One such approach is based on the Artificial Neural Network—Back Propagation (ANN-BP) model.

Objective

The objective of this research was to compare the predictive utility of the OLS-LR model with an ANN-BP approach for predicting dichotomous criteria in personnel selection problems. ANN-BP models can operate as universal approximators without *a priori* knowledge of the underlying relationships between predictors and criteria.

Although ANNs have been receiving increasing attention as models for a variety of research problems in the physical sciences, most of the ANN research in the social sciences has concentrated on models of cognition. Relatively little work has been done on the application of ANN models to prediction. This research used computer-simulated data to compare ANN-BP and OLS-LR methods for personnel selection.

Approach

The research used computer-simulated data to compare two prediction methods: OLS-LR and ANN-BP. The effects of the following dimensions were studied:

1. Function form of the data distributions
2. Base rate
3. Selection ratio
4. Validity
5. Sample size
6. Sample split

A cross-validation design was employed. Each sample was split into a development and an evaluation subsample. Both the OLS-LR and ANN-BP models were developed on the development subsample, and then tested on the corresponding evaluation subsample. The index of model performance was the "hit rate"(i.e., the proportion of correct decisions made).

Findings

For 62 comparisons there was a significant difference between the two methods ($p < .001$) in the curvilinear distributions. No significant differences were observed between the two methods in the linear distributions. In 61 of these 62 cases, the ANN-BP model outperformed the OLS-LR model. Total sample size was important, with 56 of the significant differences observed in the largest sample of 5,000, six in the samples of 500, and none with the samples of 100. There were no significant differences due to base rates, selection ratios or sample splits.

Discussion and Conclusions

This exploratory research examined the usefulness of ANN-BP for predicting dichotomous criteria in personnel selection. The results are quite encouraging. Under a wide variety of circumstances, the ANN-BP model was superior to the OLS-LR model in predicting curvilinear relationships. Only once was OLS-LR superior. When the underlying relationship was linear, there were no significant differences were found. This is somewhat remarkable, given that the OLS-LR model was designed to perform optimally in the linear case.

The results of this research indicate that additional examination of ANN-BP models in personnel selection would likely be useful.

Contents

	Page
Introduction	1
Background	1
Objective	1
Method	2
Data Generation	2
Ordinary Least-Squares—Linear Regression Models	3
Artificial Neural Network—Back Propagation Models	3
Analyses	4
Results	4
Discussion and Conclusions	4
References	7
Distribution List	9

Introduction

Background

Many of the important criteria for military personnel selection are dichotomous (or continuous variables that have been dichotomized). An example of these dichotomous criteria is successful completion of first tour of obligated service versus premature attrition. The efficacy of prediction models for forecasting these dichotomous criteria is a paramount issue in personnel selection research.

Frequently, these prediction models are developed using Ordinary Least-Squares—Linear Regression (OLS-LR). This model assumes that the underlying relationship between variables is linear. This may limit its predictive efficacy when the assumption is violated. One alternative model for predicting dichotomous criteria is the Artificial Neural Network—Back Propagation (ANN-BP) model.

The topic of Artificial Neural Networks (ANNs) has been receiving considerable attention in many basic and applied research fields (Hertz, Krogh & Palmer, 1991; McClelland & Rumelhart, 1986; Nelson & Illingworth, 1991; NeuralWare (1991a); and Rumelhart & McClelland, 1986). The breadth of this research ranges from robotics (Martinetz & Schulten,1990) to backgammon playing (Tesauro, 1990). Recently, there has been increasing interest in the application of ANN technology to research issues in the military manpower and personnel arena (Dempsey, Harris, & Waters, 1990; Sands, 1991, 1992, 1993; Sands & Wilkins, 1991; Wiggins, Looper, & Engquist, 1991a, 1991b; and Wilkins & Dickieson, 1991).

There are two general types of neural networks: biological and artificial. Stanley (1989), in an introduction to the topic, defines a neural network as "A massively parallel, dynamic system of highly interconnected, interacting parts based on neurobiological models. The behavior of the network depends heavily on the connection details" (p. 13). For this report, a more focused definition of artificial neural networks is "Computer methods which learn, through experience, the interrelationships between variables, without requiring prior information about those relationships" (Wilkins & Dickieson, 1991). The ANNs act as universal approximators. This means that they can model any relationship of interest between variables, to any desired degree of accuracy, given a large enough network. Unfortunately, there is no guarantee that the Back Propagation algorithm will successfully iterate to the desired configuration of network weights. This issue must be investigated empirically.

ANNs have been given various names in the literature, including: parallel distributed processing models, connectionist/connectionism models, adaptive systems, self-organizing systems, neurocomputing, and neuromorphic systems (Nelson & Illingworth, 1991).

Objective

The objective of this research was to compare the predictive utility for personnel selection of an OLS-LR model and an ANN-BP model. Ideally, this comparison should provide results that allow generalization to a diverse set of personnel selection situations. This would include variations in the following dimensions of the general personnel selection problem:

Base Rate—percentage of persons accepted under the current selection procedure who are successful.

Selection Ratio—percentage of persons who are selected using the new selection instrument.

Validity—the correlation of the new selection instrument with the criterion.

Sample Size—the number of cases available for statistical analyses.

Sample Split—the allocation of total sample data between a model development (or training) subsample and a model evaluation (or testing) subsample.

This goal of broad generalization suggested the use of computer-simulated data. These data would have the advantage of being "well-behaved" in a statistical sense. Use of any empirical data set runs the risk of limiting (perhaps severely) the extent to which the results may be generalized.

For most of these dimensions, no specific hypotheses were established, since the goal of the research was simply to explore what possible effects varying these dimensions might have on predictive efficacy of ANNs. However, it was hypothesized that there would be no significant differences in predictive utility between ANN-BP and OLS-LR models for the linear case. More importantly, it was hypothesized that ANN-BP would frequently predict better than OLS-LR in the curvilinear case.

Method

Data Generation

A FORTRAN computer program was developed to simulate bivariate data sets in which a continuous predictor variable and a dichotomous criterion variable had both linear and curvilinear relationships with each other. The data sets were simulated by sampling from bivariate populations in which the predictor variable was normally distributed and the criterion variable was created by dichotomizing a continuous criterion variable. For example, suppose that a continuous criterion variable called Days-In-Service represents the number of days that an applicant serves during enlistment and that Days-In-Service is used to construct a dichotomous criterion variable representing the applicants's success or failure in completing the first 2 years of enlistment (Comp-2). Then, assuming that Armed Forces Qualification Test (AFQT) percentile score and Days-In-Service have a bivariate normal distribution, a sample consisting of AFQT percentile score and Comp-2 is representative of the simulated linear data sets employed during the research.

Initially, a predictor and a criterion were computed for each simulated applicant (simulee). In the linear case, each simulee was given a randomly selected score from a standard normal distribution to represent the simulee's predictor score. The corresponding criterion score was calculated in two steps. First, the exact value on the underlying linear function line (Y_e) corresponding to the predictor was computed by multiplying the predictor (in z-score form) by the validity coefficient. The actual criterion score (Y_a) was then found by randomly selecting a value from a normal distribution with a mean equal to Y_e and a standard deviation equal to the standard error of estimate ($\sqrt{1-r^2}$). This procedure created a bivariate distribution of scores with a

correlation that was approximately equal to the desired validity. If the difference between the obtained and desired validity was more than .001, then a different random number seed was selected and the procedure was repeated until the correlation obtained was within .001 of the desired validity. Validity coefficients of .05, .25, .50, .75, and .90 were used in the research.

The curvilinear data sets were created using the same predictor values that were used in the linear case and fit to an inverted u-shaped curve (rather than to a straight line) using the following equation:

$$Y_e = -\frac{4}{9}x^2+2$$

After finding the exact criterion value, Y_e, the appropriate amount of error was introduced to the underlying function line, following the same procedure described for the linear case.

These total data sets were then split into two smaller data sets—one for model development (or training) and the other for model evaluation (or testing). Each of the simulated data files was used to create three pairs of output files, according to the following sample splits between the development and evaluation samples: 60/40, 50/50, and 20/80. The data were rank-ordered on the criterion, and then matched groups of simulees were randomly assigned to one of two groups, depending on the desired split. For instance, in the 50/50 split, the first two simulees were selected and one was randomly assigned to the development sample, the other to the evaluation sample. This procedure was followed for each pair of simulees. In the 20/80 sample, simulees were selected in groups of five, with one of the five being assigned to the development sample and the other four to the evaluation sample. Each simulee was classified as either a success or failure for each of the four base rates (.05, .25, .50, .95).

Ordinary Least-Squares—Linear Regression Models

A separate regression equation was produced for each development sample for each base rate. Then, the correlation between the actual (dichotomized) criterion and the predicted criterion was computed. Next, each regression equation was used to perform a cross-validation on the corresponding evaluation sample. At this point, the dependent variables for the regression portion of the research were estimated. First, the data were rank-ordered by the estimated criterion. Then, for each of the four selection ratios (.05, .25, .75, .90), the estimated criterion was artificially dichotomized to reflect acceptance or rejection of each simulee. Then the program calculated the number of correct acceptances, correct rejections, erroneous acceptances, and erroneous rejections, as well as the proportion of correct decisions (hit rate) for each of the selection ratios.

Artificial Neural Network—Back Propagation Models

The ANN-BP portion of the research required the specification of a particular network architecture, as well as a decision regarding the number of iterations to be used in training the network on the development samples. These decisions about the network were based on evaluating the performance of various networks on selected "toy" problems. The toy problems were created using a 50/50 split, a validity of .50, a base rate of .50, and a linear function form. This was done for all three sample sizes (100, 500, 5,000) and all four selection ratios (.05, .25, .75, .90).

Each of the toy data files was converted to an appropriate format for analysis using the NeuralWare Professional II/Plus software package (NeuralWare, 1991b). Five toy problem ANN-BPs were created. All five networks consisted of three layers: an input layer, a hidden layer, and an output layer. All of the toy networks had one input node and one output node, but the number of hidden nodes varied from one to five. Each of the toy problem development samples was trained five times on each of the networks, using the following number of training iterations: 10,000, 50,000, 100,000, 200,000, and 1,000,000. Each network was then cross-validated on the corresponding toy problem evaluation sample.

The correlation between the predicted and actual criterion values was computed. These values were tabled, analyzed, and compared across the various networks and iterations. Based upon these results, it was decided to use an ANN-BP model with one input node, three hidden nodes, and one output node. In addition, the decision was made to train these models for 100,000 iterations. Finally, the default values for the learning and momentum parameters were employed. The last step in the ANN-BP portion of the research was to calculate the same measure of effectiveness ("hit rate") for the ANN-BP models that was calculated for the OLS-LR models. The results from both ANN-BP and OLS-LR models were then tabled and compared.

Analyses

For each of the combinations of function form, sample size, sample split, base rate, selection ratio, and validity, the proportion of correct decisions (hit rate) was compared for the ANN-BP and OLS-LR models. The usual parametric test for this situation would have been a Student's *t*-test, but its use in this situation violated the required assumptions, prompting the use of nonparametric tests.

Each of the observed differences was tested for statistical significance, using either the McNemar test or the Binomial test. The McNemar test was used when the two methods classified at least 10 simulees differently. When less than 10 simulees were classified differently, it was necessary to use the Binomial test. The appropriate test was calculated in each case, and the statistical significance was assessed.

Results

The difference between methods was significant in 62 instances ($p < .001$). In 61 of these cases, ANN-BP outperformed OLS-LR; in the one remaining case, OLS-LR was superior to ANN-BP. All of these significant differences occurred in the curvilinear case. That is, none of the linear cases yielded significantly different results between ANN-BP and OLS-LR models. Fifty-six of these significant cases occurred with a sample size of 5,000, six with a sample size of 500, and none with a sample size of 100. No patterns of statistical significance were observed for sample splits, base rates, or selection ratios.

Discussion and Conclusions

The research effort examined a number of dimensions of the personnel selection problem. This discussion will focus only on basic trends in the data, due to the great number of combinations

studied. Furthermore, because of the sheer number of significance tests, the results must be considered carefully.

The research examined the simple case of a single predictor and a single criterion. Of the infinite number of possible underlying functional relationships between the predictor and criterion, this research looked at only two: a simple linear relationship and a simple curvilinear relationship.

In the linear case, linear regression actually had a great advantage over ANNs. The OLS-LR algorithm had, in essence, been told that the underlying function to try to fit was a straight line (i.e., it was constrained to arrive at the best possible answer).

Under these circumstances, ANN-BP techniques have the very difficult task of finding iteratively, without any **a priori** information, a relationship that simple linear regression will automatically fit. ANN-BP, which is not predisposed in any manner to find linear relationships, could easily fail to fit the linear relationships as accurately as the simple linear regression.

In all of the cases in which the underlying relationship was linear, there was no significant difference in the efficacy of ANN-BP and OLS-LR. This was hypothesized and is a very favorable result, considering the inherent advantage the OLS-LR model had over the ANN-BP model for situations involving linear relationships. Across all combinations of validities, sample sizes, sample splits, base rates, and selection ratios, ANN-BP did as well as OLS-LR in the linear case. Even when there was a lot of error, ANN-BP did not overfit the data any more or less than OLS-LR.

One of the biggest concerns of using ANN-BP instead of OLS-LR models for prediction is that many criteria are predicted quite well by OLS-LR models. It would not be desirable to adopt a new technology that outperforms OLS-LR with nonlinear data, but does not do as well as OLS-LR with linear data. The results of this research indicate that even when OLS-LR is the perfect model for prediction, ANN-BP develops a prediction model whose predictive efficacy is not significantly different from that of the OLS-LR model.

Of course, OLS-LR is a very quick and well understood procedure, whereas ANN-BP is an iterative method that is not as well understood. Therefore, if ANN-BP technology is to be considered as an alternative to OLS-LR technology for prediction, it must be demonstrated that it yields benefits above and beyond OLS-LR. It was hypothesized that such benefits, if they exist, would arise in the nonlinear portion of this research.

When the underlying function form was curvilinear, 62 cases produced a significant difference between ANN-BP and OLS-LR ($p < .001$). Of these 62 significant differences, 56 were observed with a total sample size of 5,000, six with a total sample size of 500, and none with a total sample size of 100. In 61 of these 62 cases, ANN-BP outperformed OLS-LR ($p < .001$), while OLS-LR was superior in the remaining case. While this ratio is quite impressive, the single case of OLS-LR leading to better prediction than ANN-BP invited further examination.

Obviously, in this one case, not only did the ANN-BP model fail to find the appropriate underlying function, but the function it did fit led to significantly worse performance than a simple linear regression line. This entire research, however, used a single network configuration that, for all the combinations of variables, was trained for exactly 100,000 iterations. The fact that this one

simple ANN-BP configuration led to better prediction performance than OLS-LR in 61 cases is quite a remarkable and robust finding. It is not surprising that no significant differences were found in the remaining cases, with a total sample size of 5,000. Quite possibly, this was due to a choice of ANN-BP configuration and stopping criterion that could be improved in one, or both, dimensions. In addition, the learning and momentum parameters could be varied, singly or in combination.

The one case in which OLS-LR outperformed ANN-BP provides some evidence supporting this possibility. When this same network was trained for an additional 100,000 iterations (i.e., for a total of 200,000 iterations), ANN-BP was significantly superior to OLS-LR ($p < .001$). It is probably unreasonable to expect a single ANN-BP configuration along with a single stopping criterion, and default settings for the learning and momentum parameters to be appropriate for all combinations addressed in this research. In more realistic setting, these various decisions would be optimized for a specific problem. In this research, only one set of parameters was chosen. Despite this, the ANN-BP models performed remarkably well across many combinations of dimensions for the personnel selection problem.

This research suggests that ANN-BP technology holds great promise for the prediction of dichotomous criteria. While much work remains to be done, this research has provided some evidence that is very promising.

References

Dempsey, J. R., Harris, D. A., & Waters, B. K. (1990). The use of artificial neural networks in military manpower modeling. *Proceedings of the 32nd Annual Conference of the Military Testing Association, 32*, 25-30.

Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the theory of neural computation.* Redwood City, CA: Addison-Wesley.

Martinetz, T., & Schulten, K. (1990). Hierarchical neural net for learning control of a robot's arm and gripper. *Proceedings of the International Joint Conference of Neural Networks*, 2, 747-752.

McClelland, J. L., & Rumelhart, D. E. (Eds.) (1986). *Parallel distributed processing: Explorations in the microstructure of cognition - Psychological and biological models* (Vol. 2). Cambridge, MA: MIT Press.

Nelson, M. C., & Illingworth, W. T. (1991). *A practical guide to neural nets*. Reading, MA: Addison-Wesley.

NeuralWare (1991a). *Neural computing*. Pittsburgh, PA: NeuralWare, Inc.

NeuralWare (1991b). *Reference guide*. Pittsburgh, PA: NeuralWare, Inc.

Rumelhart, D. E., & McClelland, J. L. (Eds.) (1986). *Parallel distributed processing: Explorations in the microstructure of cognition - Foundations* (Vol. 1). Cambridge MA: MIT Press.

Sands, W. A. (1991). Artificial neural networks: A tool for psychologists. *Proceedings of the 33rd annual conference of the Military Testing Association*, 33, 384-388.

Sands, W. A. (1992). Introduction to back propagation: An artificial neural network paradigm. *Proceedings of the 34th annual conference of the Military Testing Association, 34*, 843-848.

Sands, W. A. (1993, October). *Comparison of back propagation and regression models for personnel selection.* Paper presented at the 35th Annual Conference of the Military Testing Association, Williamsburg, VA.

Sands, W. A., & Wilkins, C. A. (1991). Artificial neural networks for personnel selection. *Proceedings of the 33rd annual conference of the Military Testing Association, 33*, 389-392.

Stanley, J. (1989). *Introduction to neural networks.* Sierra Madre, CA: California Scientific Software.

Tesauro, G. (1990). Neurogammon: A neural-network backgammon program. *Proceedings of the International Joint Conference of Neural Networks, 3*, 33-39.

Wiggins, V. L., Looper, L. T., & Engquist, S. K. (1991a). *Neural networks: A primer* (AL-TP-1991-0011). Brooks Air Force Base, TX: Human Resources Directorate, Manpower and Personnel Division, Armstrong Laboratory.

Wiggins, V. L., Looper, L. T., & Engquist, S. K. (1991b). *Neural networks and their applications to Air Force personnel modeling* (AL-TP-1991-0031). Brooks Air Force Base, TX: Human Resources Directorate, Manpower and Personnel Division, Armstrong Laboratory.

Wilkins, C. A., & Dickieson, J. L. (1991, July). *Using Neural Network Technology for Predicting Military Attrition.* Paper presented at the International Joint Conference of Neural Networks, Seattle, WA.

Distribution List

Distribution:
Chief of Naval Research (Code 342), (Code 01)
Defense Technical Information Center (DTIC) (4)

Copy to:
Chief of Naval Research (Code 04)
Pentagon Library
Director, Defense Personnel Security Research Center